

Soutenance de thèse

Maxime EL MASRI soutiendra sa thèse de doctorat préparée au sein de l'ISAE-ONERA MOIS et intitulée «*Echantillonnage préférentiel en grande dimension via des projections dans un sous-espace de petite dimension*»

Le 16 mars 2022 à 13h30, salle des thèses de l'ISAE-SUPAERO

devant le jury composé de

M. Florian SIMATOS	Professeur ISAE-SUPAERO	Directeur de thèse
M. Jérôme MORIO	Directeur de Recherche ONERA	Co-directeur de thèse
M. Jean-Michel MARIN	Professeur Université de Montpellier	Rapporteur
M. Bruno TUFFIN	Directeur de recherche INRIA	Rapporteur
M. Jean-Marc BOURINET	Maître de conférences Université Clermont Auvergne	
Mme Gersende FORT	Directrice de recherche IMT	
M. Olivier ZAHM	Chargé de recherche INRIA	

Résumé : De nombreuses disciplines scientifiques s'intéressent à l'estimation d'espérances d'une fonction d'intérêt selon une certaine loi de probabilité. Cette fonction peut être considérée comme une boîte noire, potentiellement coûteuse à évaluer. Une méthode couramment utilisée pour estimer des espérances, tout en limitant le nombre d'appels à la boîte noire, est la méthode stochastique d'échantillonnage préférentiel (Importance Sampling, IS) qui consiste à échantillonner selon une loi de probabilité auxiliaire au lieu de la loi initiale. L'estimateur d'IS est défini à partir de l'estimateur de Monte-Carlo, avec des poids d'importance, et converge presque sûrement vers l'espérance voulue, par la loi des grands nombres. Cependant, sa variance, et donc la précision de l'estimation, dépend fortement du choix de la densité auxiliaire. Une densité optimale d'échantillonnage préférentiel minimisant la variance peut être définie sur le plan théorique mais n'est pas connue en pratique. Une possibilité est alors de choisir la densité auxiliaire dans une famille paramétrique, avec laquelle il est facile de générer des échantillons, afin d'approcher la distribution optimale théorique. Des algorithmes adaptatifs (Adaptive Importance Sampling, AIS), qui estiment les paramètres de manière itérative, ont été développés pour trouver les paramètres optimaux permettant d'approcher la densité théorique visée. Mais lorsque la dimension de l'espace des paramètres augmente, l'estimation des paramètres se dégrade et les algorithmes d'AIS, et l'IS en général, deviennent inefficaces. L'estimation finale de l'espérance devient alors très imprécise, notamment du fait de l'accumulation des erreurs commises dans l'estimation de chaque paramètre. L'objectif principal de cette thèse est ainsi d'améliorer la précision de l'IS en grande dimension, en réduisant le nombre de paramètres estimés à l'aide de projections dans un sous-espace de petite dimension. Nous nous concentrons particulièrement sur la recherche de directions de projection influentes pour l'estimation de la matrice de covariance dans un cadre gaussien unimodal (où l'on met à jour le vecteur moyenne et la covariance). La première piste explorée est la projection sur le sous-espace de dimension un engendré par la moyenne optimale. Cette direction est particulièrement pertinente dans le cas d'estimation d'une probabilité d'événement rare, car la variance semble diminuer selon cette direction. La seconde proposition correspond à la projection optimale obtenue en minimisant la divergence de Kullback-Leibler avec la densité visée. Cette seconde proposition permet de projeter dans un espace de plusieurs dimensions contrairement à la première, et permet d'identifier les directions les plus influentes. Dans un premier temps, l'efficacité de ces projections est testée sur différents exemples d'estimation d'espérances en grande dimension, dans un cadre théorique n'impliquant pas d'algorithmes adaptatifs. Les simulations numériques réalisées montrent une nette amélioration de la précision de l'estimation par IS avec les deux techniques de projection sur tous les exemples considérés. Ensuite, nous proposons un couplage de ces projections avec l'algorithme d'Entropie Croisée (Cross Entropy, CE), un algorithme d'AIS destiné à l'estimation de probabilités d'événements rares. L'efficacité de ces algorithmes est vérifiée sur plusieurs cas-tests avec un faible budget de simulation. La technique basée sur la projection dans les directions optimales permet d'obtenir des estimations très précises pour des dimensions modérément grandes (plusieurs

dizaines). Le couplage avec la projection sur la moyenne reste en revanche performante dans des dimensions de quelques centaines dans la plupart des exemples. Dans tous les cas, les simulations montrent que les méthodes proposées sont plus précises que la CE classique en grande dimension avec un même budget.

Mots-clés : échantillonnage préférentiel, grande dimension, projection

Summary: Some physical systems are described by a relation $Y=f(X)$, where the input X is random and the output Y is determined by the deterministic function f . The reliability of such a system is an important question and corresponds to the issue of rare event probability estimation. Indeed, the failure of a system is a rare event which is essential to estimate. The Crude Monte Carlo method is not well adapted for this kind of estimation and some new techniques are more efficient, like sequential Monte Carlo or Importance Sampling. A current challenge in reliability analysis is rare event estimation where the input X has a high dimension. Indeed, Importance Sampling or sequential Monte Carlo are less accurate in this case. The goal of this thesis is to improve the performance of these methods in the high dimensional setting. First, using "active subspaces" to estimate integrals and to be able to sample only on a low dimensional subspace of the input space. Then it could be interesting to combine active subspaces and sampling methods. On the other hand, sensitivity analysis is looking for the most influent input dimensions on the reliability probability. Then it could be used to reduce the dimension of the sampling space, combining sensitivity analysis and sampling methods.

Keywords: importance sampling, high dimension, projection